THE UNIVERSITY OF TEXAS AT TYLER
Soules College of Business
Department of Computer Science
COSC 5371 – Data Mining
Summer 2025
Section: 2025-SUMMER7WK2-COSC-5371.460

Instructor: Nary Subramanian, Ph.D.
COB 315.11
Email (*preferred way to contact*): nsubramanian@uttyler.edu
Phone: 430-558-1330

Office Hours: By Zoom and email.

Text (Required): *Learn Data Mining Through Excel: A Step-by-Step Approach for Understanding Machine Learning Methods*, Hong Zhou, 2023, Apress Publishing, ISBN **978-1484297704**.

Catalog Description: Study of the concepts and techniques of data mining, or knowledge discovery in databases. The automated or convenient extraction of patterns representing knowledge implicitly stored in large databases, data warehouses, and other massive information repositories.

Course Description: It is often said that data is the new oil—and just like oil, data must be refined to extract valuable insights. Raw data is continuously collected from various sources, including sensors, switches, routers, computers, databases, filesystems, mobile phones, and the Internet of Things (IoT). However, this data must be mined to detect meaningful patterns and generate insights that can enhance business competitiveness. In modern organizations, structured data is typically stored in tables, such as spreadsheets and databases. One of the most widely used spreadsheet applications, Microsoft Excel, offers powerful tools not only for data manipulation but also for data mining. Many may be surprised to learn that Excel can perform sophisticated data mining tasks directly within spreadsheets. In this course, we will explore various data mining techniques using Excel. Excel is accessible to all enrolled UT Tyler students. All course materials and videos will be available on Canvas.

Grading: Grading will be based on homework assignments. Assignments are open book. Assignment submissions should be made electronically to Canvas. Late submissions will not be graded but there is **no** penalty for early submissions. There will be one assignment each week that will be available from 12pm on Sundays and will be due by 8am on Wednesdays; only in the last week you will have a *second* assignment that will appear at 8am that Wednesday due 8am that Friday. Each assignment will be worth 20 points and there will be seven assignments in all. For computing your final letter grade, the lowest assignment grade will be dropped and the sum of the six best assignment grades (maximum possible score = 120) will be used as per the grading policy given below.

Grading Policy:

Points	Grade
≥105	A
≥90, < 105	B
≥80, < 90	C

Course Objectives:

1. Explain data mining and its importance for modern businesses.
2. Compare data mining algorithms.
3. Determine the appropriate algorithm for a given problem.
4. Perform data mining using Excel.

Census and Drop Dates
Census date: June 27th, Drop date: July 24th.

Make-up Policy
There will be no make-ups for missed homework assignments; missed assignments will get a grade of zero.

Tentative Schedule:

Week	Chapter	Topic
6/23 – 6/29	1,2	Excel and Data Mining, Linear Regression
6/30 – 7/6	3,6	K-Means Clustering, Logistic Regression
7/7 – 7/13	7,8	K-Nearest Neighbors, Hierarchical Clustering
7/14 – 7/20	10	Decision Trees
7/21 – 7/27	11	EDA
7/28 – 8/3	13	Artificial Neural Networks
8/4 – 8/9	14	Text Mining